

10. Storage




EBS vs EFS vs S3 vs Glacier

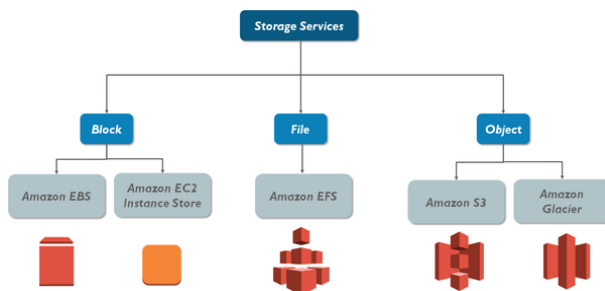
EBS: lida com volumes, podendo ser do tipo SSD ou HDD





EFS: seria um servidor de arquivos de rede e é mais utilizado com Linux. É possível mapear/montar o diretório para várias instancias. Não há limites para o tamanho de armazenamento

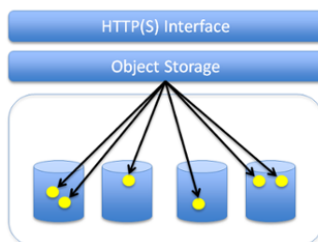
S3: storage de objetos utilizado apenas para upload e download dos objetos armazenados, diferentemente de um volume EBS em que é possível manipular os objetos dentro dele

Glacier: storage de objetos utilizado para upload e download dos objetos armazenados de longa duração, porém o download não é em tempo real. Os objetos ficam armazenados em fitas magnéticas

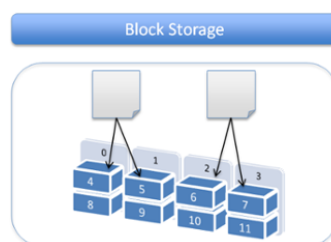
		
Amazon S3	Amazon EBS	Amazon EFS
Data storage for unstructured data	System storage for Amazon EC2 VMs	Scalable data storage for Amazon EC2 VMs



				
	Amazon EBS	Amazon S3	Amazon Glacier	Amazon EFS
Interface	Block	HTTP(S)/API	API	NFSv4
Concurrency	Single interface	Many connections	Low	Thousands
Consistency	Strongly consistent	Eventually consistent	Consistent	Strongly consistent
Latency	Low latency	Higher latency	3 options available	Low latency
Size	16 TiB/volume	5 TiB/item, unlimited	40 TiB/archive, unlimited	47.9 TiB/file
Durability	0.1-0.2% AFR AZ-specific volumes	99.999999999%	99.999999999%	Multi-AZ storage
Use Case	DBs, file servers	Static storage, high concurrency	Archival	Shared files, Big data analysis



- Store virtually unlimited files.
- Maintain file revisions.
- HTTP(S) based interface.
- Files are distributed in different physical nodes.



- File is split and stored in fixed sized blocks.
- Capacity can be increased by adding more nodes.
- Suitable for applications which require high IOPS, database, transactional data.

Amazon Elastic Block Store (Amazon EBS)

O Amazon Elastic Block Store (EBS) é um serviço de armazenamento em bloco fácil de usar e de alta performance, projetado para uso com o Amazon Elastic Compute Cloud (EC2) para taxas de transferência e cargas de trabalho intensivas de transações em qualquer escala.

Projetados para sistemas essenciais à missão, os volumes EBS são replicados em uma zona de disponibilidade (AZ) e podem ser facilmente escalonados para petabytes de dados.

Além disso, é possível usar o EBS Snapshots com políticas de ciclo de vida automatizadas para fazer backup de seus volumes no Amazon S3 e, ao mesmo tempo, garantir a proteção geográfica de seus dados.

EBS Snapshots

Os snapshots do Amazon EBS fornecem uma solução simples e segura de proteção de dados projetada para proteger seus dados de armazenamento em bloco, como volumes EBS, volumes de inicialização, bem como dados de bloco locais.

EBS Snapshots são uma cópia pontual de seus dados e podem ser usados para habilitar a recuperação de desastres, migrar dados entre regiões e contas e melhorar a conformidade de backup. Você pode criar e gerenciar seus EBS Snapshots por meio do AWS Management Console, do AWS CLI ou dos AWS SDKs.

Volumes EC2

Dentro do EBS - Elastic Block Storage (virtual disk) existem volumes que permitem que se rode um sistema operacional e existem volumes que não permitem.

Utiliza também como unidade de medida para saber a eficácia dos discos a sigla IOPS (Input/Output Operations Per Second), HDs SSD fazem de 5000 a 100.000 IOPS.

Os tipos de volumes são:

- GP2 – utiliza discos SSD, mais caro \$, de 3 a 10.000 IOPS
- IO1 (Provisioned SSD) – Alta intensidade (voltado para banco de dados), mais de 10.000 a 20.000 IOPS
- ST1 – utiliza HDD, utilizado para dados e logs. Não Boot (não roda um S.O), não pode ser (C:)
- SC1 (Cold HDD) – dados e logs e é mais barato que o ST1. Infrequent Access, não pode ser (C:)
- Magnetic (Standard) – HDD, Infrequent Access, aceita Boot (pode ser C:)

Que volume devo escolher?

O Amazon EBS inclui duas grandes categorias de armazenamento: armazenamento baseado em SSD para cargas de trabalho transacionais (a performance depende principalmente das IOPS, da latência e da durabilidade) e armazenamento sustentado por HDD para cargas de trabalho de taxa de transferência (a performance depende principalmente da taxa de transferência, medida em MB/s).

Os volumes baseados em SSD são criados para cargas de trabalho de bancos de dados transacionais com alto consumo de IOPS, volumes de inicialização e cargas de trabalho que exigem IOPS elevadas. Os volumes baseados em SSD incluem SSD de IOPS provisionadas (io1 e io2) e SSD de uso geral (gp2).

O io2 é a geração mais recente dos volumes SSD de IOPS provisionadas projetada para fornecer 100 vezes a durabilidade de 99,999%, o que a torna ideal para aplicações essenciais para os negócios que precisam de mais tempo de atividade.

Os volumes sustentados por HDD são criados para cargas de trabalho com alto consumo da taxa de transferência e de big data, E/S extensas e padrões de E/S sequenciais. Os volumes baseados em HDD incluem Throughput Optimized HDD (ST1) e Cold HDD (SC1).

Volume Types

Amazon EBS provides the following volume types, which differ in performance characteristics and price, so that you can tailor your storage performance and cost to the needs of your applications. The volumes types fall into two categories:

- SSD-backed volumes optimized for transactional workloads involving frequent read/write operations with small I/O size, where the dominant performance attribute is IOPS.
- HDD-backed volumes optimized for large streaming workloads where throughput (measured in MiB/s) is a better performance measure than IOPS.

	Solid-State Drives (SSD)		Hard disk Drives (HDD)	
Volume Type	General Purpose SSD (gp2)*	Provisioned IOPS SSD (io1)	Throughput Optimized HDD (st1)	Cold HDD (sc1)
Description	General purpose SSD volume that balances price and performance for a wide variety of workloads	Highest-performance SSD volume for mission-critical low-latency or high-throughput workloads	Low cost HDD volume designed for frequently accessed, <u>throughput-intensive workloads</u>	Lowest cost HDD volume designed for less frequently accessed workloads
Use Cases	<ul style="list-style-type: none">• Recommended for most workloads• System boot volumes• Virtual desktops• Low-latency interactive apps• Development and test environments	<ul style="list-style-type: none">• Critical business applications that require sustained IOPS performance, or more than 10,000 IOPS or 160 MiB/s of throughput per volume• Large database workloads, such as:<ul style="list-style-type: none">◦ MongoDB◦ Cassandra◦ Microsoft SQL Server◦ MySQL◦ PostgreSQL◦ Oracle	<ul style="list-style-type: none">• Streaming workloads requiring consistent, fast throughput at a low price• <u>Big data</u>• Data warehouses• Log processing• Cannot be a boot volume	<ul style="list-style-type: none">• Throughput-oriented storage for large volumes of data that is infrequently accessed• Scenarios where the lowest storage cost is important• Cannot be a boot volume
API Name	gp2	io1	st1	sc1
Volume Size	1 GiB - 16 TiB	4 GiB - 16 TiB	500 GiB - 16 TiB	500 GiB - 16 TiB
Max. IOPS**/Volume	10,000	32,000***	500	250
Max. Throughput/Volume	160 MiB/s****	500 MiB/s†	500 MiB/s	250 MiB/s
Max. IOPS/Instance	80,000	80,000	80,000	80,000
Max. Throughput/Instance††	1,750 MiB/s	1,750 MiB/s	1,750 MiB/s	1,750 MiB/s
Dominant Performance Attribute	IOPS	IOPS	MiB/s	MiB/s

Use a Modern Linux Kernel

Use a modern Linux kernel with support for indirect descriptors. Any Linux kernel 3.11 and above has this support, as well as any current-generation EC2 instance. If your average I/O size is at or near 44 KiB, you may be using an instance or kernel without support for indirect descriptors. For information about deriving the average I/O size from Amazon CloudWatch metrics, see [I/O Characteristics and Monitoring](#).

To achieve maximum throughput on st1 or sc1 volumes, we recommend applying a value of 256 to the `xen_blkfront.max` parameter (for Linux kernel versions below 4.6) or the `xen_blkfront.max_indirect_segments` parameter (for Linux kernel version 4.6 and above). The appropriate parameter can be set in your OS boot command line.

Use RAID 0 to Maximize Utilization of Instance Resources

Some instance types can drive more I/O throughput than what you can provision for a single EBS volume. You can join multiple gp2, io1, st1, or sc1 volumes together in a RAID 0 configuration to use the available bandwidth for these instances. For more information, see [RAID Configuration on Linux](#).

Encryption Volum

Changing the Encryption State of Your Data

There is no direct way to encrypt an existing unencrypted volume, or to remove encryption from an encrypted volume. However, you can migrate data between encrypted and unencrypted volumes. You can also apply a new encryption status while copying a snapshot:

- While copying an unencrypted snapshot of an unencrypted volume, you can encrypt the copy. Volumes restored from this encrypted copy are also encrypted.
- While copying an encrypted snapshot of an encrypted volume, you can associate the copy with a different CMK. Volumes restored from the encrypted copy are only accessible using the newly applied CMK.

You cannot remove encryption from an encrypted snapshot.

Amazon Data Lifecycle Manager

Amazon Data Lifecycle Manager can be used for creation, retention & deletion of EBS snapshots. It protects critical data by initiating backup of Amazon EBS volumes at selected intervals, along with storing & deletion of old snapshots to save storage space & cost.

Links Úteis

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html>

➤ Pontos de Atenção

1. Você acionou a criação de um instantâneo de seu volume EBS e está em andamento no momento. Neste ponto, o volume pode ser usado normalmente enquanto o snapshot está em andamento.
2. Uma empresa global de apostas esportivas online tem seu popular aplicativo da web hospedado na AWS. Eles estão planejando desenvolver um novo portal online para seu novo empreendimento e contrataram você para implementar a arquitetura em nuvem para um novo portal online que aceitará apostas globalmente para esportes mundiais. Você começou a projetar o sistema com um banco de dados relacional executado em uma única instância EC2, que requer um único volume EBS que pode suportar até 30.000 IOPS. Nesse cenário, qual tipo de volume do Amazon EBS você pode usar para atender aos requisitos de desempenho deste novo portal online? R: EBS Provisioned IOPS SSD
3. Você está construindo um novo aplicativo de análise de dados na AWS que será implantado em um grupo Auto Scaling de instâncias EC2 On-Demand e um banco de dados MongoDB. Espera-se que o banco de dados tenha cargas de trabalho de alto rendimento executando pequenas operações de E / S aleatórias. Como arquiteto de soluções, você deve configurar e iniciar adequadamente os recursos necessários na AWS. Qual das opções a seguir é o tipo de EBS mais adequado para usar em seu banco de dados? R: Provisioned IOPS SSD
4. You are working for a tech company that uses a lot of EBS volumes in their EC2 instances. An incident occurred that requires you to delete the EBS volumes and then re-create them again. What step should you do before you delete the EBS volumes? R: Store a snapshot of the volume.
5. A WordPress website hosted in an EC2 instance, which has an additional EBS volume attached, was mistakenly deployed in the us-east-1a Availability Zone due to a misconfiguration in your CloudFormation template. There is a requirement to quickly rectify the issue by moving and attaching the EBS volume to a new EC2 instance in the us-east-1b Availability Zone. As the Solutions Architect of the company, which of the following should you do to solve this issue? R: First, create a snapshot of the EBS volume. Afterwards, create a volume using the snapshot in the other Availability Zone.

6. You are creating a Provisioned IOPS volume in AWS. The size of the volume is 10 GiB. Which of the following is the correct value that should be put for the IOPS of the volume? R: 500
7. You have an On-Demand EC2 instance with an attached EBS volume. There is a scheduled job that creates a snapshot of this EBS volume every midnight at 12 AM when the instance is not used. One night, there has been a production incident where you need to perform a change on both the instance and on the EBS volume at the same time, when the snapshot is currently taking place. Which of the following scenario is true when it comes to the usage of an EBS volume while the snapshot is in progress? **R: The EBS volume can be used while the snapshot is in progress.**
8. You are using AWS EC2 linux instance for log processing which would require high throughput. You chose Throughput optimized HDD storage with 500 GB in size. You deployed your application to production mode and it is running as expected. After a month, you see an increase in log files and you are fast approaching the 500 GB size and running out of space on the EBS volume. Which of the following is a best approach to mitigate the situation with minimal configuration? **R: Increase the size of the existing EBS volume.**
9. Uma empresa possui um conjunto de instâncias EC2 que hospedam aplicativos geradores de receita. Alguns dos dados nos volumes raiz do EBS são essenciais para reter. Portanto, deve-se garantir que mesmo após as instâncias serem encerradas, os volumes EBS permanecerão intactos. Qual das opções a seguir precisa ser feita para garantir que esse requisito seja atendido? **R: Make the attribute of DeleteOnTermination for the EBS volume to false.**
10. Uma empresa deseja obter redundância para seus volumes EBS hospedados na AWS da maneira mais econômica. Como conseguir isso? **R: Nothing, since by default, EBS Volumes are replicated within their Availability Zones.**

Amazon Elastic File System (Amazon EFS)

O Amazon Elastic File System (Amazon EFS) fornece um sistema de arquivos elástico simples e sem servidor para definição única que permite compartilhar dados de arquivos sem provisionar ou gerenciar o armazenamento.

Ele pode ser usado com serviços da Nuvem AWS e recursos on-premises e foi criado para dimensionamento sob demanda até petabytes sem interromper as aplicações.

Com o Amazon EFS, você pode expandir e reduzir seus sistemas de arquivos automaticamente à medida que adiciona e remove arquivos, dispensando a necessidade de provisionar e gerenciar capacidade para acomodar o crescimento.

O Amazon EFS oferece a opção de criar sistemas de arquivos usando as classes de armazenamento Standard ou One Zone. As classes de armazenamento Standard armazenam dados dentro e em várias zonas de disponibilidade (AZ).

As classes de armazenamento One Zone armazenam dados de maneira redundante em uma única AZ, a um preço 47% menor em comparação com sistemas de arquivos que usam classes de armazenamento Standard, para cargas de trabalho que não exigem resiliência Multi-AZ.

O Amazon EFS oferece quatro classes de armazenamento: duas classes de armazenamento Standard:

- Amazon EFS Standard
- Amazon EFS Standard-Infrequent Access (EFS Standard-IA)

Duas classes de armazenamento One Zone:

- Amazon EFS One Zone
- Amazon EFS One Zone-Infrequent Access (EFS One Zone-IA)

O Amazon EFS fornece acesso seguro a milhares de conexões para instâncias do Amazon EC2 e servidores on-premises, bem como serviços de computação da AWS, incluindo ECS, EKS, AWS Fargate e AWS Lambda, usando simultaneamente um modelo tradicional de permissões de arquivos, recursos de bloqueio de arquivos e estrutura hierárquica de diretórios por meio do protocolo NFSv4.

As instâncias do Amazon EC2 podem acessar seu sistema de arquivos em AZs, regiões e VPCs, enquanto os servidores locais podem acessar usando o AWS Direct Connect ou a AWS VPN.

Características

O Amazon EFS trabalha sobre o protocolo de NFS (Network File System)

Elástico: significa que não é preciso se preocupar com o tamanho que será utilizado, ele aumenta e diminui automaticamente. Paga de acordo com a utilização por GB.

Multi-AZ. Por ser extremamente redundante, é muito utilizado em sistemas clusterizados, ou seja, compartilhamento de disco com vários servidores

Trabalha sobre o protocolo NFSv4 (threads simultâneos)

Performance de IO se necessário. Por exemplo quando há 200 servidores acessando o mesmo volume, será necessário uma maior performance de IO

Para utilização é necessário criar um volume na página do serviço NFS, instalar um plugin dentro da instância EC2 e depois executar um comando para fazer o mount do volume dentro do servidor

Amazon EFS is not supported on Windows instances.

Amazon EFS Lifecycle

O gerenciamento do ciclo de vida do Amazon EFS gerencia automaticamente o armazenamento de arquivos de baixo custo para seus sistemas de arquivos. Quando habilitado, o gerenciamento do ciclo de vida migra arquivos que não

foram acessados por um determinado período de tempo para a classe de armazenamento EFS Standard – Infrequent Access (Standard-IA) ou One Zone – Infrequent Access (One Zone-IA), dependendo do seu sistema de arquivos.

Você define esse período de tempo usando uma política de ciclo de vida. O gerenciamento do ciclo de vida do Amazon EFS usa um cronômetro interno para rastrear quando um arquivo foi acessado pela última vez, e não os atributos do sistema de arquivos POSIX que podem ser visualizados publicamente.

Sempre que um arquivo no armazenamento Standard ou One Zone é acessado, o cronômetro de gerenciamento do ciclo de vida é reiniciado. Depois que o gerenciamento do ciclo de vida move um arquivo para uma das classes de armazenamento IA, o arquivo permanece lá indefinidamente.

As operações de metadados, como listar o conteúdo de um diretório, não contam como acesso ao arquivo. Durante o processo de transição do conteúdo de um arquivo para uma das classes de armazenamento IA, o arquivo é armazenado na classe de armazenamento Standard ou One Zone e cobrado nessa taxa de armazenamento.

O gerenciamento do ciclo de vida se aplica a todos os arquivos no sistema de arquivos. Você pode mover arquivos de uma das classes de armazenamento IA para a classe de armazenamento Standard ou One Zone copiando-os para outro local em seu sistema de arquivos. Se você deseja que seus arquivos permaneçam na classe de armazenamento Standard ou One Zone, desative o gerenciamento do ciclo de vida e copie seus arquivos.

Você define quando o Amazon EFS faz a transição dos arquivos para uma classe de armazenamento IA definindo uma política de ciclo de vida. Um sistema de arquivos possui uma política de ciclo de vida que se aplica a todo o sistema de arquivos.

Se um arquivo não for acessado durante o período de tempo definido pela política de ciclo de vida que você escolher, o Amazon EFS fará a transição do arquivo para a classe de armazenamento IA aplicável ao seu sistema de arquivos. Você pode especificar uma das seguintes políticas de ciclo de vida para seu sistema de arquivos Amazon EFS:

- AFTER_7_DAYS
- AFTER_14_DAYS
- AFTER_30_DAYS
- AFTER_60_DAYS
- AFTER_90_DAYS

Performance Modes

Para oferecer suporte a uma ampla variedade de cargas de trabalho de armazenamento em nuvem, o Amazon EFS oferece dois modos de desempenho. Você seleciona o modo de desempenho de um sistema de arquivos ao criá-lo. Os dois modos de desempenho não têm custos adicionais, portanto, seu sistema de arquivos Amazon EFS é cobrado e medido da mesma forma, independentemente do seu modo de desempenho. Para obter informações sobre os limites do sistema de arquivos, consulte Limites para sistemas de arquivos Amazon EFS. Nota: O modo de desempenho de um sistema de arquivos Amazon EFS não pode ser alterado após a criação do sistema de arquivos.

General Purpose Performance Mode

Recomendamos o modo de desempenho de propósito geral para a maioria dos seus sistemas de arquivos Amazon EFS. A finalidade geral é ideal para casos de uso sensíveis à latência, como ambientes de serviço da Web, sistemas de gerenciamento de conteúdo, diretórios iniciais e serviço de arquivo geral. **Se você não escolher um modo de desempenho ao criar seu sistema de arquivos, o Amazon EFS seleciona o modo de finalidade geral para você por padrão.**

Max I/O Performance Mode

Os sistemas de arquivos no modo Max I / O podem escalar para níveis mais altos de rendimento agregado e operações por segundo com uma compensação de latências ligeiramente mais altas para operações de arquivo.

Aplicativos e cargas de trabalho altamente paralelizados, como análise de big data, processamento de mídia e análise genômica, podem se beneficiar desse modo.

Choose performance mode

We recommend **General Purpose** performance mode for most file systems. **Max I/O** performance mode is optimized for applications where tens, hundreds, or thousands of EC2 instances are accessing the file system — it scales to higher levels of aggregate throughput and operations per second with a tradeoff of slightly higher latencies for file operations.

- ☒ **General Purpose**
- ☐ **Max I/O**

Throughput modes

There are two throughput modes to choose from for your file system, Bursting Throughput and Provisioned Throughput.

Com o modo **Bursting Throughput (Default)**, a taxa de transferência no Amazon EFS é dimensionada conforme o tamanho do seu sistema de arquivos na classe de armazenamento EFS Standard ou One Zone aumenta.

Com o modo **Provisioned Throughput**, você pode provisionar instantaneamente a taxa de transferência do seu sistema de arquivos (em MiB/s), independentemente da quantidade de dados armazenados.

O modo Provisioned Throughput está disponível para aplicativos com altas taxas de throughput to storage (MiB / s por TiB) ou com requisitos maiores do que os permitidos pelo modo Bursting Throughput. Por exemplo, digamos que você esteja usando o Amazon EFS para ferramentas de desenvolvimento, serviço da Web ou aplicativos de gerenciamento de conteúdo em que a quantidade de dados em seu sistema de arquivos é baixa em relação às demandas de rendimento. Seu sistema de arquivos agora pode obter os altos níveis de rendimento que seus aplicativos exigem sem ter que pagar seu sistema de arquivos.

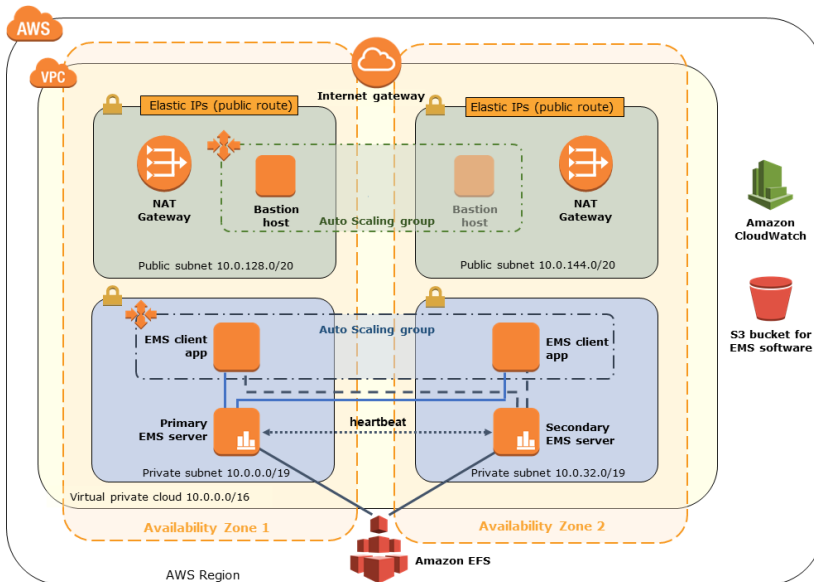
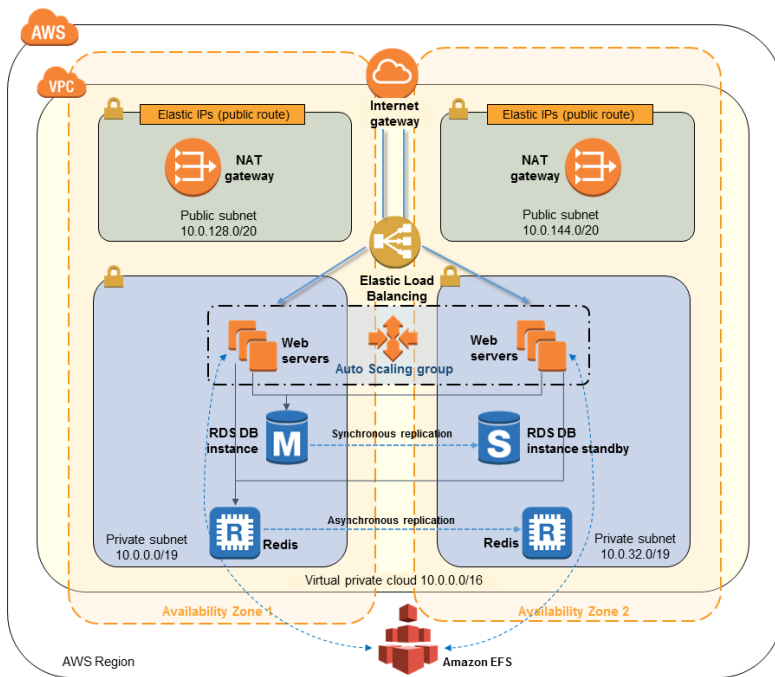
Nota: Você pode diminuir a taxa de transferência do sistema de arquivos no modo Provisioned Throughput ou alterar entre o modo Provisioned Throughput e o modo Bursting Throughput padrão, desde que tenham se passado mais de 24 horas desde a última alteração do modo de throughput ou diminuição da taxa de transferência provisionada.

EFS vs EBS

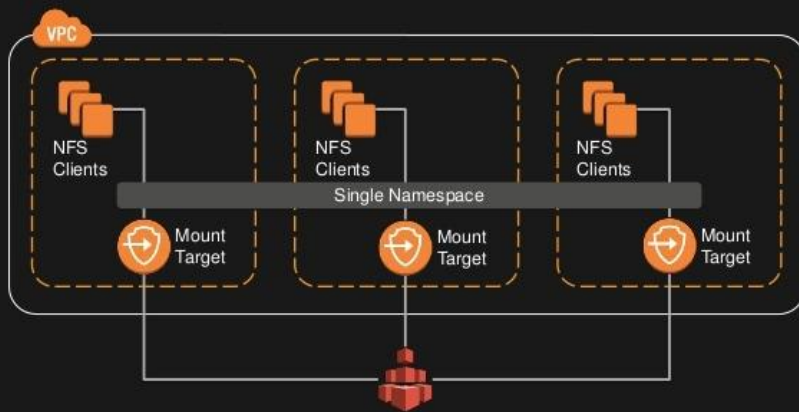
The following table shows the characteristics of EFS vs. EBS.

	Amazon EFS	Amazon EBS Provisioned IOPS
Availability and durability	Data is stored redundantly across multiple AZs.	Data is stored redundantly in a single AZ.
Access	Up to thousands of Amazon EC2 instances, from multiple AZs, can connect concurrently to a file system.	A single Amazon EC2 instance in a single AZ can connect to a file system.
Use cases	Big data and analytics, media processing workflows, content management, web serving, and home directories.	Boot volumes, transactional and NoSQL databases, data warehousing, and ETL.

Algumas Imagens



Amazon EFS Architecture



1. Uma empresa líder de comércio eletrônico precisa de uma solução de armazenamento que possa ser acessada por 1.000 servidores Linux em várias zonas de disponibilidade. O serviço deve ser capaz de lidar com os dados que mudam rapidamente em escala, mantendo o alto desempenho. Ele também deve ser altamente durável e altamente disponível sempre que os servidores extraírem dados dele, com pouca necessidade de gerenciamento. Como arquiteto de soluções, qual dos serviços a seguir é a escolha mais econômica que você deve usar para atender ao requisito acima? R: EFS
2. Você está prestes a criar um sistema de arquivos Amazon Elastic File Service (EFS) para suas instâncias EC2 e não prevê o acesso frequente a seus arquivos. Portanto, você decide escolher uma política de ciclo de vida que moverá automaticamente os arquivos para a classe de armazenamento Infrequent Access (IA) após um determinado período de tempo. Qual das opções a seguir é a política de gerenciamento de ciclo de vida mais econômica? R: **7 dias** desde o último acesso
3. A multinational company has been building its new generation big data and analytics platform in AWS in which they need a scalable storage service. The data need to be stored redundantly across multiple AZ's and allows concurrent connections from multiple EC2 instances hosted on multiple Availability Zones. Which of the following AWS storage service is the best one to use in this scenario? R: Elastic File System
4. A data analytics company has been building its new generation big data and analytics platform on their AWS cloud infrastructure. They need a storage service that provides the scale and performance that their big data applications require such as high throughput to compute nodes coupled with read-after-write consistency and low-latency file operations. In addition, their data needs to be stored redundantly across multiple AZs and allows concurrent connections from multiple EC2 instances hosted on multiple AZs. Which of the following AWS storage services will you use to meet this requirement? R: EFS
5. You have created AWS EFS with default settings and mounted on an EC2 instance. Due to regulatory policies, your organization had asked you to encrypt data stored on EFS. What would you do to enable encryption? **R: Encryption at rest option can only be set during EFS creation. You need to create encryption-at-rest EFS, copy data from old EFS to new EFS and delete old EFS.**
6. You have created AWS EFS with default settings and mounted on an EC2 instance. Due to regulatory policies, your organization had asked you to encrypt data during transit to EFS. What would you do to enable encryption during transit? **R: Enable encryption during mounting on EC2 using Amazon EFS mount helper. Unmount unencrypted mount and remount using mount helper encryption during transit option.**
7. You are building a content-serving web application with 20 EC2 instances. The EC2 servers are all load-balanced, and content storage for the instances will remain the same. You have chosen AWS EFS to act as a common storage repository. Your application needs to have as low of latency as possible when serving content to the web users. Which of the following options is the best choice for this situation? R: General Purpose Performance Mode
8. Você está construindo um aplicativo da web de veiculação de conteúdo (tamanho não superior a 25 GB) em 5 instâncias com balanceamento de carga EC2. Você escolheu o EFS para armazenamento de conteúdo. O conteúdo é acessado com frequência por um grande número de usuários. Qual dos seguintes modos de taxa de transferência você escolheria para garantir que o aplicativo em instâncias do EC2 transfira os dados para o EFS sem qualquer gargalo de desempenho? R: Throughput mode = Provisioned, you can configure specific throughput irrespective of EFS data size.
9. Você está trabalhando para um instituto financeiro usando a infraestrutura em nuvem da AWS. Todos os dados relacionados ao projeto são carregados no Amazon EFS. Esses dados são recuperados de um data center local conectado ao VPC por meio do AWS Direct Connect. Você precisa garantir que todo o acesso do cliente ao EFS seja criptografado usando TLS 1.2 para aderir às diretrizes de segurança mais recentes da equipe de segurança. Qual das opções a seguir é uma prática recomendada de baixo custo para proteger dados em trânsito ao acessar dados do Amazon EFS? **R: Use EFS mount helper to encrypt data in transit.**

Amazon FSx

Amazon FSx for Windows File Server

Armazenamento de arquivos totalmente gerenciado baseado no Windows Server

O Amazon FSx for Windows File Server fornece armazenamento de arquivos altamente confiável, escalável e totalmente gerenciado, acessível pelo protocolo SMB (Service Message Block) padrão do setor.

Ele é baseado no Windows Server, oferecendo uma ampla gama de recursos administrativos, como cotas de usuários, restauração de arquivos de usuário final e integração com o Microsoft Active Directory (AD).

Ele oferece opções de implantação Multi-AZ e single-AZ, backups totalmente gerenciados e criptografia de dados em repouso e em trânsito.

Você pode otimizar os custos e a performance das necessidades das cargas de trabalho com opções de armazenamento em SSD e HDD, e ainda dimensionar o armazenamento e alterar a performance de throughput do seu sistema de arquivos a qualquer momento.

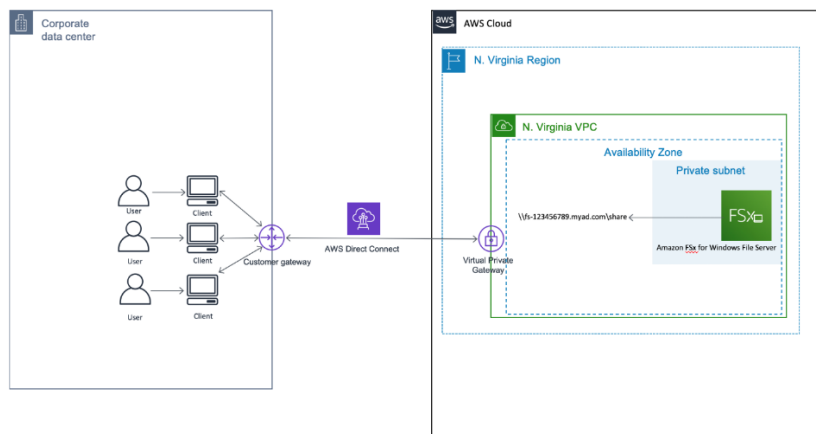
O armazenamento de arquivos do Amazon FSx pode ser acessado nas instâncias e dispositivos de computação do Windows, Linux e MacOS em execução na AWS ou no local.

O Amazon FSx integra-se ao Microsoft Active Directory (AD) on-premises e ao AWS Microsoft Managed AD. Com o suporte nativo do Windows Server para o protocolo SMB, as aplicações baseadas no Windows têm acesso a um armazenamento de arquivos compartilhado totalmente compatível.

Uma vez que o arquivo SMB também pode ser acessado pelo Linux e MacOS, qualquer aplicativo ou usuário pode acessar o armazenamento, independentemente do sistema operacional.

Características

As of 2019, Amazon FSx for Windows File Server supports access across VPCs, accounts, and Regions via Direct Connect or VPN (on-premises) and VPC Peering or AWS Transit Gateway.



FSx for Windows File Server vs FSx for Lustre

- FSx for Windows File Server is well suited to have a shared storage for your Windows instances. But it does not read data from S3, and it isn't a high-performance storage.
- **FSx for Lustre is a high-performance storage. It can read data from S3 and connect to multiple instances at the same time.**

➤ Pontos de Atenção

1. The fraud detection department in a financial analytics company using Amazon Web Services with recommended configurations needs to transfer data from their POSIX-compliant file system (Lustre) to an Amazon S3 bucket. In this context, which statement is correct? R: Amazon FSx for Lustre integrates natively with Amazon S3.
2. Você tem várias instâncias fazendo aprendizado de máquina para calcular. Você tem todos os dados necessários para o aprendizado de máquina em um intervalo S3. Você precisa encontrar um armazenamento de alto desempenho em que todas as instâncias possam ler e gravar dados simultaneamente. Qual das opções a seguir é a solução mais adequada para isso? R: FSx for Lustre.
3. Você tem um cluster de instâncias do Windows unidas a um AWS Managed Active Directory. Você deseja ter um armazenamento compartilhado para todas essas instâncias e controlar esse acesso ao armazenamento com o Active Directory gerenciado. Que serviço permite que você consiga isso? R: FSx for Windows File Server.

Amazon S3

O Amazon Simple Storage Service (Amazon S3) é um serviço de armazenamento de objetos que oferece escalabilidade líder do setor, disponibilidade de dados, segurança e performance

Isso significa que clientes de todos os tamanhos e setores podem usá-lo para armazenar qualquer volume de dados em uma grande variedade de casos de uso

Tais como sites, aplicações para dispositivos moveis, backup e restauração, arquivamento, aplicações empresariais, dispositivos IoT e análises de big data

O Amazon S3 foi projetado para 99,999999999% (11 9s) de durabilidade e armazena dados para milhões de aplicativos para empresas de todo o mundo

O Amazon S3 é altamente redundante pois os arquivos são copiados para diversas AZs justamente para garantir a segurança necessária de que os arquivos estarão disponíveis quando for preciso

Projeta os dados contra acesso não autorizado com recursos de criptografia e ferramentas de gerenciamento de acesso

O S3 mantém diversos padrões de conformidade de segurança (PCI-DSS, HIPAA/HITECH, FedRAMP, EU Data Protection Directive e FISMA) para ajudar a cumprir requisitos normativos, como por exemplo uma aplicação financeira que tem compliances de segurança a seguir

Execute análises de big data nos objetos do S3 (e em outros conjuntos de dados na AWS) com os serviços de consulta local

Use o **Amazon Athena** para consultar dados do S3 com expressões SQL padrão e o **Amazon Redshift Spectrum** para analisar dados armazenados nos data warehouses e recursos do S3 na AWS

Categorias de Armazenamento do S3

Como funciona a configuração de armazenamento do S3

As classes de armazenamento S3 podem ser configuradas no nível do objeto, em um único Bucket pode conter objetos armazenados no S3 Standard, S3 Intelligent-Tiering, S3 Standard-IA e no S3 One Zone-IA

Também podemos usar as políticas de ciclo de vida do S3 para migrar automaticamente objetos entre classes de armazenamento sem nenhuma alteração nos aplicativos

Amazon S3 Standard: Oferece um armazenamento de objetos com altos níveis de resiliência, disponibilidade e performance **para dados acessados com frequência**. Como fornece baixa latência e alto throughput (taxa de transferência), o S3 Standard é adequado para uma grande variedade de casos de uso, como aplicativos na nuvem, sites dinâmicos, distribuição de conteúdo, aplicativos moveis e de jogos analíticos de big data

Amazon S3 Standard-Infrequent Access (S3 Standard-IA): É indicado para dados acessados com **menos frequência**, mas que exige acesso rápido quando necessários. Oferece altos níveis de resiliência e throughput e a baixa latência do S3 Standard. A combinação de baixo custo e alta performance a tornaram ideal para armazenamento de longa duração, backups e datastores para arquivos de recuperação de desastres

Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA): Ao contrário de outras classes de armazenamento do S3, que armazena dados em no mínimo três Zonas de Disponibilidade (AZs), armazena dados em uma única AZ. É uma opção de menor custo para dados acessados com pouca frequência, mas não precisam de disponibilidade e da resiliência como no S3 Standard ou S3 Standard-IA

Amazon S3 Intelligent-Tiering (S3 Intelligent-Tiering): O S3 Inteligente-Hierarquização foi projetado para otimizar os custos movendo automaticamente os dados para o nível de acesso mais econômico, sem impacto na performance ou sobrecarga operacional. Ela funciona a partir do armazenamento de objetos em dois níveis de acesso: um nível é otimizado para acesso frequente e o outro nível de custo mais baixo, otimizado para acesso infrequente

Amazon S3 Glacier (Arquivamento)

É uma classe de armazenamento segura, durável e de baixo custo para arquivamento de dados. Pode-se armazenar com confiabilidade qualquer volume de dados a um custo competitivo ou inferior ao custo de soluções no local. Para manter os custos baixos, mas com condições de suprir necessidades variáveis, o S3 Glacier disponibiliza três opções de recuperação, que podem levar de alguns minutos a várias horas

Amazon S3 Glacier Deep Archive (Arquivamento): É a classe de armazenamento mais barata do Amazon S3 e oferece suporte à retenção e preservação digitais de longo prazo para dados que podem ser acessados **uma ou duas vezes** por ano. Essa classe é projetada para clientes que mantêm conjuntos de dados por 7 a 10 anos ou mais para cumprir requisitos de conformidade normativa, e podem ser restaurados em até 12 horas

Amazon S3 Reduced Redundancy Storage (RRS)

O Armazenamento de Redundância Reduzida (Dados acessados com frequência, não críticos) é uma opção de armazenamento do Amazon S3 que permite aos clientes armazenar dados reproduzíveis e que não sejam de fundamental importância com níveis de redundância mais baixos do que o armazenamento padrão do Amazon S3.

Ele disponibiliza uma solução altamente disponível para distribuir ou compartilhar conteúdo armazenado de forma durável em outro lugar ou para o armazenamento de miniaturas, mídia transcodificada ou outros dados processados que podem ser facilmente reproduzidos.

A opção RRS armazena objetos em vários dispositivos em diversas instalações, oferecendo durabilidade 400 vezes maior que a de uma unidade comum de disco, mas não replica objetos tantas vezes quanto o armazenamento padrão do Amazon S3. O armazenamento de redundância reduzida é:

1. Disponibilidade amparada pelo Acordo de Nível de Serviço do Amazon S3
2. Projetado para fornecer 99,99% de durabilidade e 99,99% de disponibilidade de objetos em um determinado ano. Esse nível de durabilidade corresponde a uma perda anual esperada média de 0,01% de objetos.
3. Projetado para sustentar a perda de dados em uma única instalação.

Resumo das Características, Padrão, Acesso Infrequente e RRS

	Padrão	Acesso infrequente Standard	Reduced Redundancy Storage
Durabilidade	99,999999999%	99,999999999%	99,99%
Disponibilidade	99,99%	99,9%	99,99%
Tolerância a falhas simultânea da instalação	2	2	1
Suporte a SSL	Sim	Sim	Sim
Latência de primeiro byte	Milissegundos	Milissegundos	Milissegundos
Políticas de gerenciamento de ciclo de vida	Sim	Sim	Sim

Resumo das Classes de Armazenamento:

	Classe de armazenamento	Projetado para	Zonas de disponibilidade	Duração mínima de armazenamento	Tamanho mínimo do objeto faturável	Taxas de monitoramento e escalonamento automático	Taxas de recuperação
<input checked="" type="radio"/>	Padrão	Dados acessados com frequência	≥ 3	-	-	-	-
<input type="radio"/>	Intelligent-Tiering	Dados de longa duração com padrões de acesso mutáveis ou desconhecidos	≥ 3	30 dias	-	Taxas aplicáveis por objeto	-
<input type="radio"/>	Padrão-IA	Dados de longa duração, acessados com pouca frequência	≥ 3	30 dias	128 KB	-	Taxas aplicáveis por GB
<input type="radio"/>	Uma zona-IA	Dados de longa duração, raramente acessados, não críticos	1	30 dias	128 KB	-	Taxas aplicáveis por GB
<input type="radio"/>	Glacier	Arquivamento de dados de longo prazo com tempos de recuperação variando entre minutos e horas	≥ 3	90 dias	-	-	Taxas aplicáveis por GB
<input type="radio"/>	Glacier Deep Archive	Arquivamento de dados de longo prazo com tempos de recuperação de até 12 horas	≥ 3	180 dias	-	-	Taxas aplicáveis por GB
<input type="radio"/>	Redundância reduzida	Dados acessados com frequência, não críticos	≥ 3	-	-	-	Taxas aplicáveis por GB

Classes de Armazenamento e Suas Características:

	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier	S3 Glacier Deep Archive**
Projetado para resiliência	99,999999999% (11 9's)	99,999999999% (11 9's)	99,999999999% (11 9's)	99,999999999% (11 9's)	99,999999999% (11 9's)	99,999999999% (11 9's)
Projetado para disponibilidade	99,99%	99,9%	99,9%	99,5%	99,99%	99,99%
Acordo de nível de serviço de disponibilidade	99,9%	99%	99%	99%	99,9%	99,9%
Zonas de disponibilidade	≥3	≥3	≥3	1	≥3	≥3
Cobrança mínima de capacidade por objeto	N/D	N/D	128 KB	128 KB	40 KB	40 KB
Cobrança mínima de duração de armazenamento	N/D	30 dias	30 dias	30 dias	90 dias	180 dias
Taxa de recuperação	N/D	N/D	por GB recuperado	por GB recuperado	por GB recuperado	por GB recuperado
Latência de primeiro byte	milissegundos	milissegundos	milissegundos	milissegundos	selecione minutos ou horas	selecione horas
Tipo de armazenamento	Objeto	Objeto	Objeto	Objeto	Objeto	Objeto
Transições de ciclo de vida	Sim	Sim	Sim	Sim	Sim	Sim

Observações: As categorias **Infrequent Access (Acesso Ocasional)** são mais baratas no armazenamento porem nas solicitações de recuperação de dados são mais caros. Reparar também que as solicitações do tipo POST são mais caras do que as solicitações do tipo GET

Nível gratuito da AWS

Como parte do [nível gratuito da AWS](#), você pode começar a usar gratuitamente o Amazon S3. Após o cadastro, os novos clientes da AWS recebem 5 GB de armazenamento padrão do Amazon S3 na classe de armazenamento S3 Standard, 20.000 solicitações GET, 2.000 solicitações PUT, COPY, POST ou LIST e 15 GB de transferência de dados para fora a cada mês, por um ano.

Outros Pontos Importantes:

Upload de Um Arquivo: Para fazer upload de um arquivo maior que 160 GB, use a AWS CLI, o SDK da AWS ou a API REST do Amazon S3. Lembrando que um único objeto pode ter o tamanho de **1 byte até 5 terabytes**.

URL Assinadas (Disponível pelo CloudFront): Um signed URL inclui informações adicionais, por exemplo, uma data e hora de expiração, que proporcionam a você mais controle sobre o acesso a seu conteúdo. Suponha que outros sites estão linkando a mesma imagem (armazenada no S3) que está disponível no seu site gerando custos adicionais pelos acessos. A solução é utilizar URLs Assinadas com data de expiração pois assim, de tempos em tempos, a URL mudará e os acessantes parasitas perderão o acesso aquele link após a expiração da URL

Registros de Log de Acesso ao Servidor: O registro de acesso ao servidor fornece registros detalhados para as solicitações feitas a um intervalo. Os logs de acesso ao servidor são úteis para muitos aplicativos. Por exemplo, as informações do log de acesso podem ser úteis em auditorias de segurança e acesso. Também pode ajudá-lo a aprender sobre sua base de clientes e entender sua fatura do Amazon S3. Quando os logs são habilitados o console do S3 atualiza automaticamente a lista de controle de acesso (ACL) do bucket. É recomendado que ao se habilitar o registro de logs, um outro bucket que não o atual seja selecionado como destino dos logs, diminuindo o faturamento de armazenamento e melhorando a localização de logs

Eventos de Dados do AWS CloudTrail: O Amazon S3 é integrado ao AWS CloudTrail, um serviço que fornece um registro das ações realizadas por um usuário, função ou serviço AWS no Amazon S3. O CloudTrail captura um subconjunto de chamadas de API para Amazon S3 como eventos, incluindo chamadas do console do Amazon S3 e chamadas de código para as APIs do Amazon S3. Se você criar uma trilha, poderá habilitar a entrega contínua de eventos CloudTrail para um bucket do Amazon S3, incluindo eventos para Amazon S3. Se você não configurar uma trilha, ainda poderá ver os eventos mais recentes no console do CloudTrail em Histórico de eventos. Usando as informações coletadas pelo CloudTrail, você pode determinar a solicitação feita ao Amazon S3, o endereço IP do qual

a solicitação foi feita, quem fez a solicitação, quando foi feita e detalhes adicionais. Para saber mais sobre o CloudTrail, incluindo como configurá-lo e habilitá-lo, consulte o AWS CloudTrail User Guide.

Amazon S3 Transfer Acceleration: é um recurso de nível de bucket que permite transferências rápidas, fáceis e seguras de arquivos em longas distâncias entre seu cliente e um bucket S3. O Transfer Acceleration tira proveito dos pontos de presença globalmente distribuídos no Amazon CloudFront. Conforme os dados chegam em um ponto de presença, eles são roteados para o Amazon S3 por um caminho de rede otimizado.

Pagamento pelo Solicitante: Em geral, os proprietários de bucket pagam por todos os custos de armazenamento e transferência de dados do Amazon S3 que estão associados a seu bucket. No entanto, você pode configurar um bucket para ser um bucket do Requester Pays. Com os buckets de Requester Pays, o solicitante, em vez do proprietário do bucket, paga o custo da solicitação e o download dos dados do bucket. O proprietário do intervalo sempre paga o custo de armazenamento de dados.

Hospedagem de Site Estático: Pode-se usar o Amazon S3 para hospedar um site estático. Em um site estático, as páginas da web individuais incluem conteúdo estático. Eles também podem conter scripts do lado do cliente. Por outro lado, um site dinâmico depende do processamento do lado do servidor, incluindo scripts do lado do servidor, como PHP, JSP ou ASP.NET. O Amazon S3 não oferece suporte a scripts do lado do servidor, mas a AWS tem outros recursos para hospedar sites dinâmicos. Para saber mais sobre hospedagem de sites na AWS, consulte Web Hosting. Importante: sites feitos em React e Angular são perfeitamente hospedados no S3 pois não há um back-end que gera as páginas e envia, tais frameworks somente chamam a API de back-end para o conteúdo

Lista de Controle de Acesso (ACL): As listas de controle de acesso (ACLs) do Amazon S3 permitem que você gerencie o acesso a **buckets e objetos**. Cada bloco e objeto tem uma ACL anexada a ele como um sub-recurso. Ele define quais contas ou grupos da AWS têm acesso concedido e o tipo de acesso. Quando uma solicitação é recebida em um recurso, o Amazon S3 verifica a ACL correspondente para verificar se o solicitante tem as permissões de acesso necessárias. Quando você cria um bucket ou um objeto, o Amazon S3 cria uma ACL padrão que concede ao proprietário do recurso controle total sobre o recurso. **IMPORTANTE: Quando você concede acesso aos favorecidos dos grupos Todos ou Usuários autenticados, qualquer pessoa no mundo pode acessar esse objeto. A ACL lista concessões, que identificam o beneficiário e a permissão concedida. Além de conceder acesso a Todos ou a Usuários Autenticados, é possível também conceder acesso a Beneficiários específicos**

Compartilhamento de Recursos de Origem Cruzada (CORS): O compartilhamento de recursos de origem cruzada (CORS) define uma maneira para os aplicativos da Web cliente que são carregados em um domínio interagirem com os recursos em um domínio diferente. Com o suporte do CORS, você pode construir aplicativos da web ricos no lado do cliente com o Amazon S3 e permitir seletivamente o acesso de origem cruzada aos recursos do Amazon S3. Esta seção mostra como habilitar o CORS usando o console do Amazon S3, a API REST do Amazon S3 e os SDKs da AWS. Para configurar seu bucket para permitir solicitações de origem cruzada, você adiciona uma configuração CORS ao bucket. Uma configuração CORS é um documento que define regras que identificam as origens que você permitirá que você acesse seu bucket, as operações (métodos HTTP) com suporte para cada origem e outras informações específicas da operação. No console S3, a configuração do CORS deve ser um documento JSON.

Regras de Ciclo de Vida: Para gerenciar seus objetos de forma que sejam armazenados de maneira econômica durante todo o seu ciclo de vida, configure o Amazon S3 Lifecycle. Uma configuração de ciclo de vida do S3 é um conjunto de regras que definem ações que o Amazon S3 aplica a um grupo de objetos. Existem dois tipos de ações: **Ações de transição** - defina quando os objetos fazem a transição para outro usando classes de armazenamento do Amazon S3. Por exemplo, você pode optar por fazer a transição de objetos para a classe de armazenamento S3 Standard-IA 30 dias depois de criá-los ou arquivar objetos para a classe de armazenamento S3 Glacier um ano após criá-los. **Ações de expiração** - defina quando os objetos expiram. O Amazon S3 exclui objetos expirados em seu nome.

Criar Regra de Replicação: A replicação permite a cópia automática e assíncrona de objetos em buckets do Amazon S3. Buckets configurados para replicação de objeto podem pertencer à mesma conta da AWS ou a contas diferentes. Os objetos podem ser replicados para um único intervalo de destino ou vários intervalos de destino. Os buckets de destino podem estar em diferentes regiões da AWS ou na mesma região do bucket de origem. **Regras: A replicação requer que o versionamento seja habilitado para o bucket de origem. É possível criar várias réplicas para buckets de**

destino distintos para um único bucket de origem. Uma função do IAM fará a gestão das políticas. É possível alterar a classe de armazenamento do bucket de destino dos objetos replicados

Versionamento do Bucket: O controle de versão no Amazon S3 é um meio de manter várias variantes de um objeto no mesmo intervalo. Você pode usar o recurso S3 Versioning para preservar, recuperar e restaurar todas as versões de todos os objetos armazenados em seus depósitos. Com o controle de versão, você pode se recuperar mais facilmente de ações não intencionais do usuário e de falhas do aplicativo. Depois que o controle de versão é habilitado para um bucket, se o Amazon S3 receber várias solicitações de gravação para o mesmo objeto simultaneamente, ele armazena todos esses objetos. Ao trabalhar com o controle de versão S3 em buckets do Amazon S3, você pode opcionalmente adicionar outra camada de segurança configurando um bucket para habilitar a exclusão de MFA (autenticação multifator). Ao fazer isso, o proprietário do intervalo deve incluir duas formas de autenticação em qualquer solicitação para excluir uma versão ou alterar o estado de controle de versão do intervalo. **IMPORTANTE:** A cada nova versão do mesmo arquivo sendo versionado dentro do S3 o tamanho total de todas as versões é somando, podendo aumentar o custo de armazenamento. O que pode resolver o problema são os LifeCycles com ação e expiração, onde o arquivo será removido quando expirado

Criptografia Padrão: Com a criptografia padrão do Amazon S3, você pode definir o comportamento de criptografia padrão para um bucket do S3 para que todos os novos objetos sejam criptografados quando são armazenados no bucket. Os objetos são criptografados usando criptografia do lado do servidor com chaves gerenciadas pelo Amazon S3 (SSE-S3) ou chaves AWS KMS armazenadas no AWS Key Management Service (AWS KMS) (SSE-KMS). Ao configurar seu bucket para usar criptografia padrão com SSE-KMS, você também pode habilitar S3 Bucket Keys para diminuir o tráfego de solicitação do Amazon S3 para o AWS Key Management Service (AWS KMS) e reduzir o custo da criptografia. Quando você usa criptografia do lado do servidor, o Amazon S3 criptografa um objeto antes de salvá-lo no disco e o descriptografa quando você baixa os objetos. **SSE-S3:** A chave de criptografia é criada e gerenciada pelo Amazon S3. A criptografia do lado do servidor do Amazon S3 usa uma das cifras de bloco mais fortes disponíveis para criptografar seus dados, padrão de criptografia avançada de 256 bits (AES-256). Não há novos encargos para o uso de criptografia do lado do servidor com chaves gerenciadas pelo Amazon S3 (SSE-S3). No entanto, as solicitações para configurar e usar SSE-S3 incorrem em cobranças de solicitação padrão do Amazon S3. **SSE-KMS:** O Amazon S3 usa as Chaves Mestras do Cliente (CMKs) AWS KMS para criptografar seus objetos do Amazon S3. O AWS KMS criptografa apenas os dados do objeto. Todos os metadados do objeto não são criptografados. Se você usar CMKs, use o AWS KMS por meio do AWS Management Console ou AWS KMS APIs para criar centralmente CMKs, definir as políticas que controlam como os CMKs podem ser usados e auditar seu uso para provar que estão sendo usados corretamente. Você pode usar esses CMKs para proteger seus dados em buckets do Amazon S3. Ao usar a criptografia SSE-KMS com um intervalo S3, o AWS KMS CMK deve estar na mesma região do intervalo.

SSE-S3 requires that Amazon S3 manage the data and master encryption keys.

SSE-C requires that you manage the encryption key.

SSE-KMS requires that AWS manage the data key, but you manage the master key in AWS KMS.

Bloqueio de Objetos (Object Lock): Com o S3 Object Lock, você pode armazenar objetos usando um modelo write-once-read-many (WORM). O Bloqueio de Objeto pode ajudar a evitar que objetos sejam excluídos ou sobrescritos por um período de tempo fixo ou indefinidamente. Você pode usar o Object Lock para ajudar a atender aos requisitos regulamentares que exigem armazenamento WORM ou simplesmente adicionar outra camada de proteção contra alterações e exclusão de objetos. O Bloqueio de objeto oferece duas maneiras de gerenciar a retenção de objetos: períodos de retenção e retenções legais. **Período de retenção** - especifica um período fixo de tempo durante o qual um objeto permanece bloqueado. Durante esse período, seu objeto fica protegido por WORM e não pode ser sobrescrito ou excluído. **Retenção legal** - oferece a mesma proteção que um período de retenção, mas não tem data de expiração. Em vez disso, uma retenção legal permanece em vigor até que você a remova explicitamente. As retenções legais são independentes dos períodos de retenção.

Utilizando URL Pré Assinada

Tem por objetivo permitir o acesso externo a um objeto do bucket mesmo quando a Lista de Controle de Acesso (ACL) está com status ativado/bloqueado. Após o tempo de expiração, o objeto volta a se tornar restrito

A pre-signed URL gives you access to the object identified in the URL, provided that the creator of the pre-signed URL has permissions to access that object. **That is, if you receive a pre-signed URL to upload an object, you can upload the object only if the creator of the pre-signed URL has the necessary permissions to upload that object.**

All objects and buckets by default are private. The pre-signed URLs are useful if you want your user/customer to upload a specific object to your bucket, but you don't require them to have AWS security credentials or permissions. When you create a pre-signed URL, you must provide your security credentials and then specify a bucket name, an object key, an HTTP method (PUT for uploading objects), and an expiration date and time. The pre-signed URLs are valid only for the specified duration.

Algumas Características

1. Amazon S3 does not support object locking for concurrent writers.

Amazon S3 does not support object locking for concurrent writers. If two PUT requests are simultaneously made to the same key, the request with the latest timestamp wins. If this is an issue, you will need to build an object-locking mechanism into your application

2. Any read that is initiated following the receipt of a successful PUT response will return the data written by the PUT

Amazon S3 achieves high availability by replicating data across multiple servers within AWS data centers. If a PUT request is successful, your data is safely stored. Any read (GET or LIST) that is initiated following the receipt of a successful PUT response will return the data written by the PUT

3. Updates to a single key are atomic

Updates to a single key are atomic. For example, if you PUT to an existing key from one thread and perform a GET on the same key from a second thread concurrently, you will get either the old data or the new data, but never partial or corrupt data.

Request Rate and Performance Guidelines

Amazon S3 automatically scales to high request rates. For example, your application can achieve at least 3,500 PUT/POST/DELETE and 5,500 GET requests per second per prefix in a bucket. There are no limits to the number of prefixes in a bucket. It is simple to increase your read or write performance exponentially. For example, if you create 10 prefixes in an Amazon S3 bucket to parallelize reads, you could scale your read performance to 55,000 read requests per second.

Cross-region

A replicação entre regiões é uma configuração em nível de bucket que permite a cópia automática e assíncrona de objetos em buckets em diferentes regiões da AWS. Referimo-nos a esses intervalos como intervalo de origem e intervalo de destino. Esses depósitos podem pertencer a diferentes contas da AWS.

The source and destination buckets must have versioning enabled

Requirements

Requirements for cross-region replication:

- The source and destination buckets must have versioning enabled. For more information about versioning, see [Using Versioning](#).
- The source and destination buckets must be in different AWS Regions. For a list of AWS Regions where you can create a bucket, see [Regions and Endpoints](#) in the *AWS General Reference*.
- Amazon S3 must have permissions to replicate objects from that source bucket to the destination bucket on your behalf.

You can grant these permissions by creating an IAM role. For more information about IAM roles, see [Create an IAM Role](#).

Important

To pass the IAM role that you create that grants Amazon S3 replication permissions, you must have the `iam:PassRole` permission. For more information, see [Granting a User Permissions to Pass a Role to an AWS Service](#) in the *IAM User Guide*.

- If the source bucket owner also owns the object, the bucket owner has full permissions to replicate the object. If not, the object owner must grant the bucket owner the `READ` and `READ_ACP` permissions via the object ACL. For more information about Amazon S3 actions, see [Specifying Permissions in a Policy](#). For more information about resources and ownership, see [Amazon S3 Resources](#).

System-defined object metadata

Date

Content-Length

Content-Type

Last-Modified

Content-MD5

x-amz-server-side-encryption

x-amz-version-id

x-amz-delete-marker

x-amz-storage-class

x-amz-website-redirect-location

x-amz-server-side-encryption-aws-kms-key-id

x-amz-server-side-encryption-customer-algorithm

Links Uteis

https://docs.aws.amazon.com/pt_br/lambda/latest/dg/with-s3.html

https://docs.aws.amazon.com/pt_br/AmazonCloudFront/latest/DeveloperGuide/private-content-creating-signed-url-canned-policy.html

<https://docs.aws.amazon.com/AWSJavaScriptSDK/latest/AWS/S3.html#getSignedUrl-property>

https://docs.aws.amazon.com/pt_br/AmazonS3/latest/userguide/WebsiteHosting.html

<https://docs.aws.amazon.com/AmazonS3/latest/dev/cloudtrail-logging.html#cloudtrail-logging-vs-server-logs>

<https://docs.aws.amazon.com/AmazonS3/latest/userguide/LogFormat.html>

<https://docs.aws.amazon.com/AmazonS3/latest/dev/ServerLogs.html>

https://aws.amazon.com/s3/features/#Query_in_Place

<https://docs.aws.amazon.com/AmazonS3/latest/dev/UsingEncryption.html>

<https://docs.aws.amazon.com/AmazonS3/latest/dev/crr.html>

<https://docs.aws.amazon.com/AmazonS3/latest/dev/PresignedUrlUploadObject.html>

<https://docs.aws.amazon.com/AmazonS3/latest/userguide/UsingMetadata.html#object-metadata>

➤ Pontos de Atenção

1. Um site de compartilhamento de documentos está usando a AWS como sua infraestrutura de nuvem. Os usuários gratuitos podem fazer upload de um total de 5 GB de dados, enquanto os usuários premium podem fazer upload de até 5 TB. Seu aplicativo carrega os arquivos do usuário, que podem ter um tamanho máximo de arquivo de 1 TB, para um S3 Bucket. Nesse cenário, qual é a melhor maneira do aplicativo carregar os arquivos grandes no S3? **R: Use Multipart Upload**
2. Você está trabalhando como consultor de TI para uma grande empresa de mídia onde tem a tarefa de projetar um aplicativo da web que armazena ativos estáticos em um bucket do Amazon Simple Storage Service (S3). Você espera que este bucket S3 receba imediatamente mais de 2.000 solicitações PUT e 3.500 solicitações GET por segundo no horário de pico. O que você deve fazer para garantir o desempenho ideal? **R: Fazer nada. O Amazon S3 gerenciará automaticamente o desempenho nessa escala.**
3. Um de seus clientes está aproveitando o Amazon S3 na região ap-sudeste-1 para armazenar seus vídeos de treinamento para o processo de integração de seus funcionários. O cliente está armazenando os vídeos usando a classe Standard Storage. Onde os vídeos de treinamento do seu cliente são replicados? **R: Várias instalações em ap-sudeste-1**
4. Você está trabalhando como arquiteto de soluções para uma empresa financeira multinacional. Eles têm uma plataforma de comércio online global na qual os usuários de todo o mundo carregam regularmente terabytes de dados transacionais para um balde S3 centralizado. Qual recurso da AWS você deve usar em seu sistema atual para melhorar o rendimento e garantir a transferência de dados consistentemente rápida para o bucket do Amazon S3, independentemente da localização do seu usuário? **R: Amazon S3 Transfer Acceleration**
5. Sua empresa possui um aplicativo de e-commerce que salva os logs de transações em um bucket do S3. Você é instruído pelo CTO a configurar o aplicativo para manter os logs de transações por um mês para fins de solução de problemas e, em seguida, limpar os logs. O que você deve fazer para cumprir esse requisito? **R: Configure as regras de configuração do ciclo de vida no bucket do Amazon S3 para limpar os logs de transação após um mês**
6. Ao configurar as propriedades de um intervalo S3, qual das opções a seguir você deve selecionar para rastrear o custo de armazenamento? **R: Tags**
7. Você é um arquiteto de soluções em sua empresa e trabalha com três engenheiros de DevOps abaixo de você. Um dos engenheiros excluiu acidentalmente um arquivo hospedado no Amazon S3, o que causou a interrupção do serviço. O que você pode fazer para evitar que isso aconteça novamente? **R: Enable S3 Versioning and Multi-Factor Authentication Delete on the bucket.**
8. Seu cliente tem clientes em todo o mundo que acessam arquivos de produtos armazenados em vários buckets S3, que estão por trás de cada uma de suas próprias distribuições da web do CloudFront. Atualmente, eles desejam entregar seu conteúdo a um cliente específico e precisam ter certeza de que apenas esse cliente pode acessar os dados. Atualmente, todos os seus clientes podem acessar seus buckets S3 diretamente usando um URL S3 ou por meio de sua distribuição do CloudFront. Quais das opções a seguir são possíveis soluções que você poderia implementar para atender aos requisitos acima? **R: Use URLs pré-assinados S3 para garantir que apenas**

seus clientes possam acessar os arquivos. Remova a permissão para usar URLs do Amazon S3 para ler os arquivos para qualquer outra pessoa.

9. You have launched a travel photo sharing website using Amazon S3 to serve high-quality photos to visitors of your website. After a few days, you found out that there are other travel websites linking and using your photos. This resulted in financial losses for your business. What is an effective method to mitigate this issue? **R: Configure your S3 bucket to remove public read access and use pre-signed URLs with expiry dates.**
10. An application which uses multiple EBS volumes could not cope with the growing storage requirements needed to store their data. Your IT Manager has instructed you to set up an S3 bucket as a replacement for their EBS volumes. Which of the following options is correct regarding the naming convention for the S3 bucket? **R: A bucket name must be unique across all existing bucket names in Amazon S3.**
11. A startup is building an AI-based face recognition application in AWS, where they store millions of images in an S3 bucket. As the Solutions Architect, you have to ensure that each and every image uploaded to their system is stored without any issues. What is the correct indication that an object was successfully stored when you put objects in Amazon S3? **R: HTTP 200 result code and MD5 checksum.**
12. You are employed by a large electronics company that uses Amazon Simple Storage Service. For reporting purposes, they want to track and log every request access to their S3 buckets including the requester, bucket name, request time, request action, referrer, turnaround time, and error code information. The solution should also provide more visibility into the object-level operations of the bucket. Which is the best solution among the following options that can satisfy the requirement? **R: Enable server access logging for all required Amazon S3 buckets.**
13. A real-time data analytics application is using AWS Lambda to process data and store results in JSON format to an S3 bucket. To speed up the existing workflow, you have to use a service where you can run sophisticated Big Data analytics on your data without moving them into a separate analytics system. Which of the following group of services can you use to meet this requirement? **R: S3 Select, Amazon Athena, Amazon Redshift Spectrum**
14. You are working for a major financial firm in Wall Street where you are tasked to design an application architecture for their online trading platform which should have high availability and fault tolerance. The application is using an Amazon S3 bucket located in the us-east-1 region to store large amounts of intraday financial data. To avoid any costly service disruptions, what will you do to ensure that the stored financial data in the S3 bucket would not be affected even if there is an outage in one of the Availability Zones or a regional service failure in us-east-1? **R: Enable Cross-Region Replication.**
15. A Fortune 500 company which has numerous offices and customers around the globe has hired you as their Principal Architect. You have staff and customers that upload gigabytes to terabytes of data to a centralized S3 bucket from the regional data centers, across continents, all over the world on a regular basis. At the end of the financial year, there are thousands of data being uploaded to the central S3 bucket which is in ap-southeast-2 (Sydney) region and a lot of employees are starting to complain about the slow upload times. You were instructed by the CTO to resolve this issue as soon as possible to avoid any delays in processing their global end of financial year (EOFY) reports. Which feature in Amazon S3 enables fast, easy, and secure transfer of your files over long distances between your client and your Amazon S3 bucket? **R: Transfer Acceleration**
16. Em uma agência governamental para a qual você está trabalhando, você foi designado para colocar documentos fiscais confidenciais na nuvem da AWS. No entanto, há uma preocupação do ponto de vista da segurança sobre o que pode ser colocado na AWS. Quais são os recursos da AWS que podem garantir a segurança dos dados para seus documentos confidenciais? (Escolha 2) **R: 1) S3 Server-Side Encryption 2) S3 Client-Side Encryption**

17. You are trying to enable Cross-Region Replication to your S3 bucket but this option is disabled. Which of the following options is a valid reason for this? **R: In order to use the Cross-Region Replication feature in S3, you need to first enable versioning on the bucket.**
18. Sua organização precisa atender à conformidade de auditoria e, portanto, precisa registrar todas as solicitações enviadas para 10 depósitos que contêm informações confidenciais. Eles também serão usados periodicamente para determinar se alguma solicitação está sendo feita de fora do intervalo de endereços IP da organização. Sua equipe de aplicativos AWS habilitou o registro de acesso ao servidor S3 por meio do AWS Console para todos os depósitos em um depósito de registro comum denominado s3-server-logging. Mas depois de algumas horas, eles notaram que nenhum registro estava sendo gravado no depósito de registro. Qual seria a razão? **R: A política de negação definida pelo usuário do intervalo não permite que o grupo de entrega de log grave no intervalo de registro S3**
19. You created a bucket named “myfirstwhizbucket” in the US West region. What are valid URLs for accessing the bucket? (Choose 2 options) **R: 1) <https://myfirstwhizbucket.s3.us-west-1.amazonaws.com> 2) <https://s3.us-west-1.amazonaws.com/myfirstwhizbucket>**
20. You have a version enabled S3 bucket. You have accidentally deleted an object which contains 3 versions. You would want to restore the deleted object. What can be done? **R: Delete the delete-marker on the object**
21. O que a replicação entre regiões requer que o controle de versão esteja habilitado? **R: Both Source and Destination buckets**
22. You have created an S3 bucket in the us-east-1 region with default configuration. Versioning is not enabled. You are located in Asia and deleted an object in the bucket using AWS CLI. What may happen when you try to list the objects in the bucket? **R: The object is deleted**
23. What are the minimum and maximum file sizes that can be stored in S3 respectively? **R: 0 Bytes and 5 terabytes**
24. Your company requires that S3 objects should be replicated in different AWS regions. You have an S3 bucket in the ap-southeast-1 region, and its objects are encrypted with AWS Key Management Service (AWS KMS). How would you configure the Cross-Region Replication (CRR) for the encrypted objects in the S3 bucket? **R: In the replication rule, provide the KMS key name for decrypting source objects. Because users can choose one or more KMS keys in the replication rule**
25. You have an S3 bucket that is used to store important data for a web application. You want to receive an email notification whenever an object removal event happens in the S3 bucket. How would you configure the S3 bucket to achieve this requirement? **R: Configure an S3 event notification for the object removal events. Send the events to an SNS topic.**
26. Um grande instituto educacional está usando intervalos do Amazon S3 para salvar dados de todos os fluxos de formatura. Durante as auditorias externas anuais de órgãos do governo local, os institutos precisam buscar dados de fluxos específicos para que sejam auditados pelos auditores. Uma grande quantidade de dados é salva nesses depósitos S3, tornando complicado fazer o download de dados inteiros e recuperar apenas uma pequena quantidade de informações deles. A equipe de TI está procurando sua consultoria para esse problema sem incorrer em custos adicionais ou comprometer a segurança. Qual das seguintes ações é recomendada para resolução? **R: Store objects in JSON format compressing it with GZIP using server-side encryption. Use Amazon S3 Select to retrieve a subset of data. Amazon S3 Select can be used to query a subset of data from the objects stored in the S3 bucket using simple SQL. For using this, objects need to be stored in an S3 bucket with CSV, JSON, or Apache Parquet format. GZIP & BZIP2 compression is supported with CSV or JSON format with server-side encryption.**

27. Uma empresa está planejando armazenar documentos confidenciais em um balde S3. Eles querem garantir que os documentos sejam criptografados em repouso. Eles desejam garantir o gerenciamento das chaves subjacentes usadas para criptografia, mas não do processo de criptografia/descriptografia. Qual das opções a seguir pode ser usada para esse propósito? **R: Use S3 server-side encryption with Customer keys.**
28. Você está desenvolvendo um aplicativo usando o AWS SDK para obter objetos do AWS S3. Os objetos têm tamanhos grandes. Às vezes, ocorrem falhas ao obter objetos, especialmente quando a conectividade da rede é ruim. Você deseja obter um intervalo específico de bytes em uma única solicitação GET e recuperar o objeto inteiro em partes. Qual método pode conseguir isso? **R: Use the "Range" HTTP header in a GET request to download the specified range bytes of an object.**
29. Você criou um bucket S3 na região us-east-1 com configuração padrão. O controle de versão não está habilitado. Você está localizado na Ásia e excluiu um objeto do intervalo usando o AWS CLI. O que pode acontecer quando você tenta listar os objetos no balde? **R: The object is deleted completely.**

Amazon S3 Glacier

Classes de armazenamento de objetos de longo prazo, seguras e resilientes do Amazon S3 a partir de 1 USD por terabyte por mês. Use o Amazon S3 Glacier caso o armazenamento de baixo custo seja primordial e não seja necessário acessar os dados em milissegundos.

Amazon S3 Glacier

É uma classe de armazenamento segura, durável e de baixo custo para arquivamento de dados. Pode-se armazenar com confiabilidade qualquer volume de dados a um custo competitivo ou inferior ao custo de soluções no local. Para manter os custos baixos, mas com condições de suprir necessidades variáveis, o S3 Glacier disponibiliza **três opções** de recuperação, que podem levar de alguns minutos a várias horas

Recuperações Aceleradas (Expedited): geralmente retornam dados em 1 a 5 minutos e são excelentes para casos de uso de arquivamento ativo.

Recuperações Padrão (Standard): são concluídas em 3 a 5 horas e funcionam bem para atividades em que o tempo não é tão crucial, como dados de backup, edição de mídia ou análises de longo prazo.

Recuperações em Massa (Bulk): são a opção de recuperação mais barata, e retornam grandes quantidades de dados em 5 a 12 horas.

Amazon S3 Glacier Deep Archive (Arquivamento)

É a classe de armazenamento mais barata do Amazon S3 e oferece suporte à retenção e preservação digitais de longo prazo para dados que podem ser acessados **uma ou duas vezes** por ano. Essa classe é projetada para clientes que mantêm conjuntos de dados por 7 a 10 anos ou mais para cumprir requisitos de conformidade normativa, e oferece **duas opções** de acesso, de 12 a 48 horas

Recuperação Padrão (Standard): até 12 horas para esta velocidade de recuperação

Bulk Retrieval (Recuperação em Massa): retornará os dados em até 48 horas

Conceitos Importantes

Cofre (Vault)

Um cofre (vault) é um contêiner para armazenamento de arquivos. Um arquivo é qualquer objeto, como uma foto, vídeo ou documento, que você armazena em um cofre. Os cofres permitem que você organize seus arquivos e defina políticas de acesso e políticas de notificação.

Um único arquivo pode ter até 40 terabytes. Você pode armazenar um número ilimitado de arquivos e uma quantidade ilimitada de dados no S3 Glacier. Cada arquivo é atribuído a um ID de arquivo único no momento da criação, e o conteúdo do arquivo é imutável, o que significa que depois que um arquivo é criado, ele não pode ser atualizado.

Disparar Notificação

Você pode configurar cofres para enviar notificações para você ou seu aplicativo sempre que determinadas tarefas do S3 Glacier forem concluídas como a recuperação de arquivos e a recuperação de inventário do cofre. As notificações são entregues usando o Amazon Simple Notification Service (Amazon SNS).

Você especifica o tópico Amazon SNS a ser usado para notificações de conclusão de trabalho usando o Amazon Resource Name (ARN) do tópico. Em seguida, você seleciona quais tipos de trabalhos do S3 Glacier podem acionar o envio de uma notificação após a conclusão do trabalho. Os aplicativos ou usuários que assinam o tópico Amazon SNS recebem uma mensagem de notificação quando um trabalho do tipo que você seleciona é concluído.

Política de Recuperação

Usando as políticas de recuperação de dados do S3 Glacier, você pode gerenciar os custos de recuperação definindo limites nas atividades de recuperação em sua conta da AWS em cada região. As políticas de recuperação se aplicam

às recuperações padrão. Sendo três políticas: 1. Free Tier, 2. Taxa máxima de recuperação (GB/hora) e 3. Nenhum limite de recuperação. Nota: As políticas de recuperação de dados governam todas as atividades de recuperação em uma região.

Provisionar Capacidade

A capacidade provisionada deve ser usada quando você precisar de uma garantia de que as recuperações aceleradas estarão disponíveis quando necessário. Depois de adquirir a capacidade provisionada, todas as recuperações aceleradas que você fizer serão atendidas por sua capacidade provisionada. **Cada capacidade unidade custa 100 USD/mês**

AWS CloudTrail

As classes de armazenamento Amazon S3 Glacier e S3 Glacier Deep Archive oferecem integração avançada com o AWS CloudTrail para registrar em log, monitorar e reter atividades de chamadas da API de armazenamento para fins de auditoria.

Criptografia e Conformidade

Além disso, oferecem suporte a três formas diferentes de criptografia. Essas classes de armazenamento também oferecem suporte a padrões de segurança e certificações de conformidade. Além disso, o Amazon S3 Object Lock habilita recursos de armazenamento WORM, o que ajuda a cumprir os requisitos de conformidade de praticamente todos os órgãos normativos do mundo

Definição de Preço

Como parte do nível de uso gratuito da AWS, você pode recuperar gratuitamente até 10 GB de dados do Amazon S3 Glacier por mês. A franquia do nível gratuito pode ser usada a qualquer momento durante o mês e é aplicável às recuperações **padrão**. O uso mensal não utilizado não é acumulado para o mês seguinte.

Links Úteis

<https://docs.aws.amazon.com/AmazonS3/latest/userguide/object-lifecycle-mgmt.html>

➤ **Pontos de Atenção**

1. Uma empresa tem 10 TB de arquivos de dados financeiros raramente acessados que precisam ser armazenados na AWS. Esses dados seriam acessados com pouca frequência durante semanas específicas, quando são recuperados para fins de auditoria. O tempo de recuperação não é restrito, desde que não exceda 24 horas. Qual das opções a seguir seria uma solução segura, durável e econômica para este cenário? **R: Faça upload dos dados para o S3 e defina uma política de ciclo de vida para fazer a transição dos dados para o Glacier após 0 dias.**
2. Your company would like to store their old yet confidential corporate files that are infrequently accessed. What cost-efficient solution in AWS should you recommend? **R: Amazon Glacier**
3. You are a Cloud Migration Engineer in a media company which uses EC2, ELB, and S3 for its video-sharing portal for filmmakers. They are using a standard S3 storage class to store all high-quality videos that are frequently accessed only during the first three months of posting. What should you do if the company needs to automatically transfer or archive media data from an S3 bucket to Glacier? **R: Use Lifecycle Policies**
4. Uma grande empresa de varejo salva seus relatórios de vendas globais no intervalo S3 e usa as regras de ciclo de vida S3 para mover esses dados da classe de armazenamento Standard_IA para AWS S3 Glacier após 180 dias. Devido ao final do ano financeiro, a equipe de Finanças está procurando um relatório de vendas apenas para a região da Europa, onde uma incompatibilidade é relatada no número de vendas. Qual das opções a seguir é uma maneira recomendada de buscar esses dados com o mínimo esforço? **R: Use Amazon S3 Glacier Select to query specific continent data directly from Amazon S3 Glacier using simple SQL. Amazon S3 Glacier Select can be used**

to query specific data from Amazon S3 Glacier instead of querying whole data. Amazon S3 Glacier Select can directly query data from Amazon S3 Glacier & restoration of data to the S3 bucket is not required for querying this data.

AWS Storage Gateway (nuvem híbrida)

O AWS Storage Gateway é um conjunto de serviços de nuvem híbrida que oferece acesso on-premises a armazenamento na nuvem praticamente ilimitado.

Os clientes usam o Storage Gateway para integrar o armazenamento da Nuvem AWS com workloads locais para que possam simplificar o gerenciamento do armazenamento e reduzir os custos de casos de uso de armazenamento fundamentais na nuvem híbrida.

Esses casos de uso incluem mover os backups para a nuvem, usar compartilhamentos de arquivo on-premises com respaldo do armazenamento na nuvem e fornecer acesso de baixa latência aos dados na AWS para aplicações on-premises.

Trocando em miúdos, um software é instalado no Data Center ou Servidor privado, e esse software terá acesso a conta da AWS e criará uma ponte entre os drivers de armazenamento privado e o bucket S3 por exemplo, para que haja a réplica dos dados locais para o bucket S3.

As aplicações se conectam ao serviço por meio de uma máquina virtual ou um dispositivo de hardware de gateway usando protocolos de armazenamento padrão, como NFS, SMB e iSCSI. O gateway se conecta a serviços de armazenamento da AWS, como o Amazon S3, Amazon S3 Glacier, Amazon S3 Glacier Deep Archive, Amazon FSx for Windows File Server, Amazon EBS e AWS Backup, fornecendo armazenamento e gerenciamento de dados para arquivos, volumes, snapshots e fitas virtuais na AWS.

O serviço inclui um mecanismo de transferência de dados otimizado e eficiente, com gerenciamento de largura de banda e resiliência de rede automatizada, além de fornecer uma experiência de gerenciamento consistente usando o Console AWS on-premises e na Nuvem AWS.

Existem 4 tipos diferentes de Gateway (com objetivos distintos), sendo eles:

Amazon S3 File Gateway (Gateway de Arquivos): Armazene e acesse objetos (vídeos, imagens, docs) no Amazon S3 a partir de dados de arquivo NFS ou SMB com cache local

Amazon FSx File Gateway: Acesso local rápido e de baixa latência para compartilhamentos de arquivos totalmente gerenciados. O Amazon FSx File Gateway otimiza o acesso local a compartilhamentos de arquivos totalmente gerenciados e altamente confiáveis no Amazon FSx para Windows File Server

Volume Gateway: Armazenamento de blocos em nuvem híbrida com cache local, normalmente direcionado para armazenamento de Sistemas Operacionais e VMWare. O Volume Gateway apresenta volumes de armazenamento de blocos iSCSI com suporte na nuvem para seus aplicativos locais. O Volume Gateway armazena e gerencia dados locais no Amazon S3 em seu nome e opera no modo de cache ou no modo armazenado. No modo Volume Gateway em cache, seus dados primários são armazenados no Amazon S3, enquanto ele retém os dados acessados com frequência localmente no cache para acesso de baixa latência. No modo Volume Gateway armazenado, seus dados principais são armazenados localmente e todo o conjunto de dados está disponível para acesso de baixa latência local, enquanto também é copiado para backup assincronamente no Amazon S3. Em ambos os modos, você pode fazer cópias pontuais dos seus volumes usando o AWS Backup, que são armazenadas na AWS como snapshots do Amazon EBS. O uso de snapshots do Amazon EBS permite que você faça cópias com controle de versão e eficiência de espaço dos seus volumes para proteção de dados, recuperação, migração e várias outras necessidades de dados de cópia.

Tape Gateway (Gateway de Fita): O serviço AWS Storage Gateway pode ser configurado para atuar como uma Virtual Tape Library (VTL – Biblioteca de fitas virtuais) que se inicia em seu ambiente local, onde estão seus aplicativos de produção, e se estende até os serviços de armazenamento redundantes, duráveis e altamente escaláveis da Nuvem AWS, o Amazon S3, o Amazon S3 Glacier e o Amazon S3 Glacier Deep Archive

Pré-requisitos

Você deve concluir as etapas a seguir antes de criar seu Amazon S3 File Gateway.

1. Crie pelo menos um bucket no Amazon S3 (Simple Storage Service).
2. Crie uma função IAM com acesso total ao gateway de armazenamento e acesso à lista, leitura e gravação S3.
3. Certifique-se de que os requisitos de porta para o seu Amazon S3 File Gateway sejam atendidos.
4. Se você estiver planejando configurar compartilhamentos de arquivos SMB com autenticação do Active Directory, adquira credenciais para um usuário administrador que pode adicionar hosts ao Active Directory associados aos usuários que acessam compartilhamentos de arquivos SMB.

Gateway de Arquivo

Para usar um gateway de arquivo, primeiro faça download de uma imagem da VM para o gateway de arquivo. Em seguida, ative o gateway de arquivos a partir do AWS Management Console ou por meio da API do Storage Gateway. Você também pode criar um gateway de arquivos usando uma imagem do Amazon EC2.

Depois que o gateway de arquivos for ativado, crie e configure o compartilhamento de arquivos e associe-o ao bucket do Amazon Simple Storage Service (Amazon S3). Ao fazer isso, o compartilhamento pode ser acessado por clientes que usam o protocolo Network File System (NFS) ou Server Message Block (SMB).

Os arquivos gravados em um compartilhamento de arquivos tornam-se objetos no Amazon S3, e o caminho funciona como chave. Não há um mapeamento individualizado entre arquivos e objetos, e o gateway atualiza assincronamente os objetos no Amazon S3 com o mesmo procedimento usado em alterações nos arquivos.

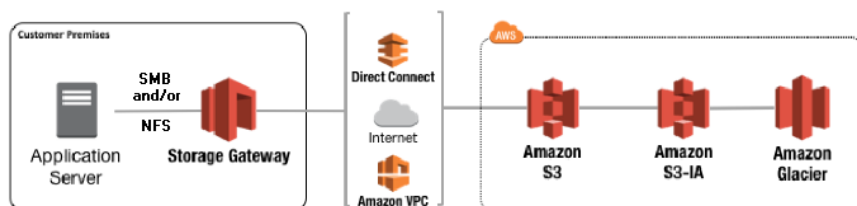
Os objetos existentes no bucket do Amazon S3 aparecem como arquivos no sistema de arquivos e a chave transforma-se o caminho. Os objetos são criptografados com chaves de criptografia no lado do servidor do Amazon S3 (SSE-S3). Todas as transferências de dados são feitas por meio de HTTPS.

O serviço otimiza a transferência de dados entre o gateway e o AWS ao usar multipart uploads paralelos ou downloads em intervalo de bytes para usar mais adequadamente a largura de banda disponível.

O cache local é mantido para fornecer acesso de baixa latência aos dados acessados recentemente e reduzir os encargos de saída de dados.

As métricas do CloudWatch fornecem informações sobre o uso de recursos na VM e a transferência de dados para e do AWS. CloudTrail rastreia todas as chamadas da API.

Com o armazenamento de gateway de arquivos, você pode realizar tarefas como ingerir cargas de trabalho da nuvem para o Amazon S3, fazer backup e arquivamento, estratificação e migração de dados de armazenamento na nuvem da Amazon Web Services. O diagrama a seguir fornece uma visão geral da implantação do armazenamento para o Storage Gateway.



O gateway de arquivos converte arquivos em objetos do S3 ao fazer upload de arquivos para o Amazon S3. A interação entre as operações de arquivo executadas em compartilhamentos de arquivos no gateway de arquivos e objetos S3 exige que determinadas operações sejam cuidadosamente consideradas ao converter entre arquivos e objetos.

Operações comuns de arquivo alteram metadados de arquivo, o que resulta na exclusão do objeto S3 atual e na criação de um novo objeto S3. A tabela a seguir mostra operações de arquivo de exemplo e o impacto em objetos do S3.

Operação de arquivos	Impacto do objeto S3	Implicação da classe de armazenamento
----------------------	----------------------	---------------------------------------

Renomear arquivo	Substitui o objeto S3 existente e cria um novo objeto S3 para cada arquivo	Taxas de exclusão antecipada e taxas de recuperação podem ser aplicadas
Renomear pasta	Substitui todos os objetos existentes do S3 e cria novos objetos do S3 para cada pasta e arquivos na estrutura da pasta	Taxas de exclusão antecipada e taxas de recuperação podem ser aplicadas
Alterar permissões de arquivo/pasta	Substitui o objeto S3 existente e cria um novo objeto S3 para cada arquivo ou pasta	Taxas de exclusão antecipada e taxas de recuperação podem ser aplicadas
Alterar a propriedade do arquivo/pasta	Substitui o objeto S3 existente e cria um novo objeto S3 para cada arquivo ou pasta	Taxas de exclusão antecipada e taxas de recuperação podem ser aplicadas
Acrescentar a um arquivo	Substitui o objeto S3 existente e cria um novo objeto S3 para cada arquivo	Taxas de exclusão antecipada e taxas de recuperação podem ser aplicadas

PARA CONTINUAR ACESSAR ARQUIVO: AWS Storage Gateway Guia do Usuário e o Link em Links Úteis

Cached Volumes vs Storage Volumes

1. Uma empresa tem muitos dados hospedados em sua infraestrutura local. Ficando sem espaço de armazenamento, a empresa quer uma solução de ganho rápido usando AWS. Deve haver baixa latência para os dados acessados com frequência. Qual das opções a seguir permitiria a fácil extensão de sua infraestrutura de dados para a AWS? R: The company could start using Gateway Cached Volumes.

Volume Gateways and Cached Volumes can be used to start storing data in S3. AWS Documentation mentions the following: You store your data in Amazon Simple Storage Service (Amazon S3) and retain a copy of frequently accessed data subsets locally. Cached volumes offer substantial cost savings on primary storage and minimize the need to scale your storage on-premises. You also retain low-latency access to your frequently accessed data.

A questão afirma que eles estão ficando sem espaço de armazenamento e precisam de uma solução para armazenar dados com AWS, em vez de um backup. Para esse propósito, os volumes em cache do gateway são apropriados, o que os ajudará a evitar o dimensionamento de seu data center local e armazenamento no serviço de armazenamento da AWS, tendo os arquivos mais recentes disponíveis para eles em baixa latência.

This is the difference between Cached and stored volumes:

Cached volumes (Acesso Frequente) – Você armazena seus dados no S3 e mantém uma cópia de subconjuntos de dados acessados com frequência localmente. Os volumes em cache oferecem economias de custo substanciais no armazenamento primário e "minimizam a necessidade de dimensionar seu armazenamento local. Você também mantém o acesso de baixa latência aos dados acessados com frequência".

Stored volumes (backup) - Se você precisar de acesso de baixa latência a todo o seu conjunto de dados, primeiro configure seu gateway local para armazenar todos os seus dados localmente. Em seguida, faça backup de instantâneos point-in-time desses dados de forma assíncrona para o Amazon S3. "Esta configuração fornece backups externos duráveis e baratos que você pode recuperar em seu data center local ou no Amazon EC2." Por exemplo, se você precisar de capacidade de substituição para recuperação de desastres, pode recuperar os backups para o Amazon EC2.

Conforme descrito na resposta: A empresa deseja uma solução de ganho rápido para armazenar dados com AWS, evitando dimensionar a configuração local em vez de fazer backup dos dados.

Na pergunta, eles mencionaram que "Uma empresa tem muitos dados hospedados em sua infraestrutura local." De infraestrutura local à infraestrutura em nuvem, você pode usar gateways de armazenamento da AWS.

Definição de Preço

Com o AWS Storage Gateway, você paga apenas pelo que usa e é cobrado de acordo com o tipo e a quantidade do armazenamento utilizado, a quantidade de solicitações efetuadas e o volume de dados transferidos para fora da AWS.

Links Úteis

https://docs.aws.amazon.com/pt_br/storagegateway/latest/userguide/StorageGatewayConcepts.html

➤ Pontos de Atenção

You are working for a tech company which currently has an on-premises infrastructure. They are currently running low on storage and want to have the ability to extend their storage using AWS cloud. Which AWS service can help you achieve this requirement? R: Amazon Storage Gateway

Algumas Imagens Auto Explicativas

